

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 00:16:21 ; Search time 11792 Seconds  
(without alignments)  
2983.250 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 MAVGSLVHVRVTLILLWFG.....LFCLHVLFKRTSKEDREG 726

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlp  
-O=/cn2 1/USPTO spool p/US10633202/runat 14062005 141725 13089/app query.fasta\_1.903  
-DB=GenEmbl -QFMT=fastgap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=25 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10633202 @CGN 1 1 5600 @runat 14062005 141725 13089 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4030	100.0	2181	6	BD086459 SVPH1-26
2	4030	100.0	2810	9	BC025378 Homo sapi
3	4030	100.0	3230	6	AX704773 Sequence
4	4030	100.0	3314	9	AF158643 Homo sapi

C	5	4030	100.0	209483	9	CNS050TGV	AL357153 Human chr
	6	4004	99.4	2426	6	A99212	A99212 Sequence 13
	7	4004	99.4	2426	6	BD196410	BD196410 Novel met
	8	4004	99.4	2442	6	AX775081	AX775081 Sequence
	9	4004	99.4	2442	9	AF029899	AF029899 Homo sapi
	10	2435	60.4	2169	6	BD1337083	BD1337083 Testis-sp
	11	2435	60.4	2169	6	AR256989	AR256989 Sequence
	12	2435	60.4	3557	9	AF158644	AF158644 Homo sapi
	13	2425.5	60.2	18756	9	CNS060C84	AL391223 Human chr
	14	2341.5	58.1	2016	6	CQ736276	CQ736276 Sequence
	15	2330	57.8	2622	2	AF158640	AF158640 Homo sapi
	16	2319	57.5	146055	2	AC009507	AC009507 Homo sapi
	17	2319	57.5	206256	9	CNS018GT	AL160191 Human Chr
	18	2108.5	52.3	91995	9	AC105914	AC105914 Homo sapi
	19	2102.5	52.2	2301	6	BD267491	BD267491 New metal
	20	2102.5	52.2	2301	9	AF171930	AF171930 Homo sapi
	21	2102.5	52.2	2364	6	BD267492	BD267492 New metal
	22	2102.5	52.2	2364	9	AF171931	AF171931 Homo sapi
	23	2102.5	52.2	2463	6	BD267493	BD267493 New metal
	24	2102.5	52.2	2463	9	AF171929	AF171929 Homo sapi
	25	2102.5	52.2	2958	9	AF134708	AF134708 Homo sapi
C	26	2053	50.9	234832	2	AC127883	AC127883 Rattus no
	27	2050	50.9	2190	10	AF251559	AF251559 Mus muscu
C	28	2050	50.9	174472	2	AC074356	AC074356 Mus muscu
	29	1994.5	49.5	2640	10	AF167403	AF167403 Mus muscu
C	30	1982.5	49.2	190094	2	AC131819	AC131819 Rattus no
C	31	1982.5	49.2	224808	2	AC094882	AC094882 Rattus no
	32	1965	48.8	168871	2	AC016117	AC016117 Homo sapi
	33	1965	48.8	178517	9	AP006246	AP006246 Homo sapi
	34	1952	48.4	4593	9	AF158639	AF158639 Homo sapi
	35	1951	48.4	154065	2	AC084079	AC084079 Homo sapi
	36	1951	48.4	154539	2	AC079187	AC079187 Homo sapi
	37	1951	48.4	186218	9	AP006248	AP006248 Homo sapi
	38	1948.5	48.3	198057	2	AC126072	AC126072 Rattus no
	39	1948.5	48.3	327299	9	AC096502	AC096502 Rattus no
	40	1945	48.3	100000	9	AB020862	AB020862 Homo sapi
	41	1940	48.1	193426	2	AC117729	AC117729 Mus muscu
	42	1934	48.0	2726	10	AY190759	AY190759 Mus muscu
C	43	1902.5	47.2	190187	2	AC121140	AC121140 Mus muscu
	44	1892.5	47.0	2806	10	AF167402	AF167402 Mus muscu
C	45	1846	45.8	256368	2	AC122753	AC122753 Mus muscu

#### ALIGNMENTS

RESULT 1	BD086459	BD086459	2181 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	SVPH1-26	DNA and polypeptide.				
DEFINITION	BD086459					
ACCESSION	BD086459.1	GI:22632069				
VERSION	JP 2001521742-A/1.					
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 2181)					
AUTHORS	Cerretti, D.P.					
TITLE	SVPH1-26 DNA and polypeptide					
JOURNAL	Patent: JP 2001521742-A 1 13-NOV-2001;					
COMMENT	IMMUNEX CORP					
	OS Homo sapiens (human)					
	PN JP 2001521742-A/1					
	PD 13-NOV-2001					
	PF 30-OCT-1998 JP 2000519084					
	PR 30-OCT-1997 US 60/063571					
	PI DOUGLAS PAT CERRETTI					
	PC C12N15/09,C07K16/40,C12N1/19,C12N1/21,C12N5/10,C12N9/64,C12Q1/37,C12Q1/68,					
	PC G01N33/68,C12N15/00,C12N5/00					
	CC SVPH1-26 DNA and polypeptide					
	FT Key location/Qualifiers					
	FT source 1. .2181					

```
FT
FEATURES
  source
    Location/Qualifiers
      1..2181
        /organism="Homo sapiens (human)"
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
  Pred. No.: 0 Length: 2181
  Score: 4030.00 Matches: 726
  Percent Similarity: 100.00% Conservative: 0
  Best Local Similarity: 100.00% Mismatches: 0
  Query Match: 100.00% Indels: 0
  DB: 6 Gaps: 0
US-10-633-202-2 (1-726) x BD086459 (1-2181)
Qy 1 MetAlaValGlyGluProLeuValHisIleArgValThrLeuLeuLeuLeuTrpPheGly 20
Db 1 ATGCGAGTGGGTGAGCCCTGGTGCACATCAGGGTCACTCTCTGCTGCTCTGGTTGGG 60
Qy 21 MetPheLeuSerIleSerGlyHisSerGlnAlaArgProSerGlnTyrPheThrSerPro 40
Db 61 ATGTTTTTGTCTATTTCTGGCCACTCTCAGCCAGGCCCTCCCACTATTTTCACTTCTCCA 120
Qy 41 GluValValIleProLeuValIleSerArgGlyArgGlyAlaIleValProGlyTyr 60
Db 121 GAAGTGGTATTCCTTTGAAGGTGATCAGCAGGGGCGAGGTGCAAGGCTCTCTGGATGG 180
Qy 61 LeuSerTyrSerLeuArgPheGlyGlyGlnArgTyrIleValHisMetArgValAsnLys 80
Db 181 CTCTCCTATAGCCTGCGGTTTGGGGGACAGAGATACATTGTCCACATGAGGGTAAATAAG 240
Qy 81 LeuLeuPheAlaAlaHisLeuProValPheThrTyrThrGluGlnHisAlaLeuLeuGln 100
Db 241 CTGTGTGTGTGTCACACCTTCTGTGTTCACTTACACAGAGCAGCATGCGCTCCAG 300
Qy 101 AspGlnProPheIleGlnAspAspCysTyrTyrHisGlyTyrValGluGlyValProGlu 120
Db 301 GATCAGCCCTTCATCCAGGATGACTGCTTACTACCATGTTATGTGGAGGGGCTCCCTGAG 360
Qy 121 SerLeuValAlaLeuSerThrCysSerGlyGlyPheLeuGlyMetLeuGlnIleAsnAsp 140
Db 361 TCCTTGTGTGGCTTAGTACCTGTTCTGGGGCTTCTTGGAAATGCTACAGATAAATGAC 420
Qy 141 LeuValTyrGluIleLysProIleSerValSerAlaThrPheGluHisLeuValTyrLys 160
Db 421 CTTGTTTTATGAAATCAAGCCAAATAGTGTCTGCCACATTTGAAACACCTAGTATATAAG 480
Qy 161 IleAspSerAspAspThrGlnPheProProMetArgCysGlyLeuThrGluGluLysIle 180
Db 481 ATAGACAGTATGATACACAGTTTCCACCTATGAGATGTGGTTAAACAGAGAAATA 540
Qy 181 AlaHisGlnMetGluLeuGlnLeuSerTyrAsnPheThrLeuLysGlnSerSerPheVal 200
Db 541 GCACACAGATGGAGTTGCAATTCATATAAATTTCACTCTGAGCAAAAGTCTCTTTTGTG 600
Qy 201 GlyTyrTrpThrHisGlnArgPheValGluLeuValValValAspAsnIleArgTyr 220
Db 601 GGCTGGTGGACCCATCAGCGTTTGTGGAGTGTGTAGTGTGGTGGTGAATAATATAGATAT 660
Qy 221 LeuPheSerGlnSerAsnAlaThrThrValGlnHisGluValPheAsnValValAsnIle 240
Db 661 CTTTTCTCTCAAGTATAGCAACACAGTCAGCATGAAGTATTTAACTGTTGCAATATA 720
Qy 241 ValAspSerPheTyrHisProLeuGluValAspValIleLeuThrGlyIleAspIleTyr 260
Db 721 GTGGATTCCTTCTATCATCTTTGGAGGTTGATGTAATTTTACTGCAATTTGATATATGG 780
Qy 261 ThrIleSerAsnProLeuProThrSerGlyAspLeuAspAsnValLeuGluAspPheSer 280
Db 781 ACTGCATCAATCCACTTCTTACCAGTGGAGACCTAGATAATGTTTTTAGAGGACTTTTCT
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 00:22:31 ; Search time 8209 Seconds  
(without alignments)  
3366.386 Million cell updates/sec

Title: US-10-633-202-2  
Perfect score: 4030  
Sequence: 1 MAVGPELVHVRVTLTLLWFG.....LFLHLVLFKRTSKEDREG 726

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cn2 1/USPTO spool p/US10633202/runat 14062005 141725 13099/app\_query.fasta\_1.903  
-DB=EST -QFMT=fasta -SUFFIX=rst -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10633202 @CGN 1.1 5180 @runat 14062005 141725 13099 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hic:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_gsei:  
9: gb\_gsei2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079.5	51.6	3245	3 BC036070	BC036070 Homo sapi
2	2050	50.9	2169	9 AY405095	AY405095 Mus muscu
3	2050	50.9	2912	3 AK014827	AK014827 Mus muscu
4	1985.5	49.3	2615	3 AK016581	AK016581 Mus muscu
5	1899.5	47.1	2871	3 AK029528	AK029528 Mus muscu
6	1781.5	44.2	2551	3 AK029590	AK029590 Mus muscu
7	1713.5	42.5	2490	3 AK029756	AK029756 Mus muscu
8	1329.5	33.0	3881	3 BC047156	BC047156 Mus muscu
9	1303	32.3	3511	3 AK077058	AK077058 Mus muscu

10	1260	31.3	3128	3 BC047939	BC047939 Homo sapi
11	1169.5	29.0	2164	3 AK029742	AK029742 Mus muscu
12	1160.5	28.8	2492	3 AK076647	AK076647 Mus muscu
13	1137.5	28.2	3174	3 AK016558	AK016558 Mus muscu
14	1125	27.9	2546	3 AK076580	AK076580 Mus muscu
15	1042	25.9	602	7 CV024284	CV024284 1534 Full
16	971	24.1	3108	3 AF326918	AF326918 Homo sapi
17	948.5	23.5	2169	9 AY405093	AY405093 Homo sapi
18	870.5	21.6	859	3 CR804792	CR804792 GR0AAA26A
19	866.5	21.5	2547	3 AK016550	AK016550 Mus muscu
20	829.5	20.6	2515	3 AK016574	AK016574 Mus muscu
21	829.5	20.6	2547	3 AK029519	AK029519 Mus muscu
22	827	20.5	2531	3 AK039384	AK039384 Mus muscu
23	825	20.5	2576	3 AK048901	AK048901 Mus muscu
24	804	20.0	804	9 CR815043	CR815043 GR0AAA4BC
25	799.5	19.8	670	9 CE337023	CE337023 tigt-g88-
26	798.5	19.8	788	8 BH363602	BH363602 CH230-509
27	792	19.7	2235	3 AK005759	AK005759 Mus muscu
28	779.5	19.3	2364	3 AK076633	AK076633 Mus muscu
c	779	19.3	461	8 AQ060341	AQ060341 CIT-HSP-2
30	777	19.3	772	9 CR815044	CR815044 GR0AAA4BC
31	776	19.3	2546	3 BC025386	BC025386 Homo sapi
32	768.5	19.1	3036	3 AK089086	AK089086 Mus muscu
33	767.5	19.0	2277	3 AK089169	AK089169 Mus muscu
34	755.5	18.7	2891	3 AK046677	AK046677 Mus muscu
c	743	18.4	631	1 AA400496	AA400496 zu70505.8
36	736	18.3	2229	3 AK048367	AK048367 Mus muscu
37	735.5	18.3	3433	3 AK029301	AK029301 Mus muscu
38	734.5	18.2	3493	3 AK051780	AK051780 Mus muscu
39	722.5	17.9	2661	9 AV399360	AV399360 Homo sapi
40	719.5	17.9	3577	3 AK053825	AK053825 Mus muscu
41	719	17.8	708	9 CR829312	CR829312 GR0AAA60D
42	709.5	17.6	2664	9 AY399362	AY399362 Mus muscu
43	696.5	17.3	1019	7 CN647602	CN647602 ILLUMIGEN
44	695.5	17.3	2607	9 AY399361	AY399361 Pan trogl
45	689.5	17.1	2250	9 AY418590	AY418590 Homo sapi

## ALIGNMENTS

RESULT 1	BC036070	3245 bp	mRNA	linear	HTC 20-SEP-2002
LOCUS	BC036070	Homo sapiens, clone IMAGE:5298677,	mRNA.		
DEFINITION	BC036070				
ACCESSION	BC036070.1	GI:23243314			
VERSION	HTC.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				

REFERENCE  
1 (bases 1 to 3245)  
Strausberg, R.  
Direct Submission  
Submitted (31-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 04:15:12 ; Search time 5730 Seconds  
(without alignments)  
786.506 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 MAVGPELVHVRVTLNLFNG.....LPCLHVLFKRTKSKDEEG 726

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 6054689 seqs, 3103772919 residues

Total number of hits satisfying chosen parameters: 12109378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO spool\_p/US10633202/runat\_14062005\_141727\_13196/app query.fasta\_1.903  
-DB=Published Applications NA\_QWFAST-FASTA SUPPFI=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bts -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10633202 @CNG 1.1 723 @runat\_14062005\_141727\_13196  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCk=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:  
22: /cgn2\_6/ptodata/2/pubpna/US10J\_NEW\_PUB.seq:  
23: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq:  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4030	100.0	2181	18	US-10-633-202-1	Sequence 1, Appli
2	4030	100.0	2578	21	US-10-956-157-1123	Sequence 1123, Ap
3	4030	100.0	3230	10	US-09-978-418-5	Sequence 5, Appli
4	4030	100.0	3230	21	US-10-485-231-5	Sequence 5, Appli
5	4004	99.4	2442	19	US-10-283-975A-397	Sequence 397, App
6	4004	99.4	2442	21	US-10-451-997-105	Sequence 105, App
7	2435	60.4	2169	14	US-10-265-125-1	Sequence 1, Appli
8	2435	60.4	2406	21	US-10-956-157-1122	Sequence 1122, Ap
9	2108.5	52.3	2958	21	US-10-956-157-1891	Sequence 1891, Ap
10	2102.5	52.2	2301	18	US-10-664-456-7	Sequence 7, Appli
11	2102.5	52.2	2364	18	US-10-664-456-8	Sequence 8, Appli
12	2102.5	52.2	2463	18	US-10-664-456-9	Sequence 9, Appli
13	2102.5	52.2	2629	17	US-10-382-600-12	Sequence 12, Appli
14	2102.5	52.2	4007	17	US-10-382-600-17	Sequence 20, Appli
15	2102.5	52.2	5112	17	US-10-382-600-2	Sequence 17, Appli
16	2102.5	52.2	5165	17	US-10-382-600-2	Sequence 2, Appli
17	2102.5	52.2	6106	17	US-10-382-600-14	Sequence 14, Appli
18	2102.5	52.2	6159	17	US-10-382-600-1	Sequence 1, Appli
19	2102.5	52.2	6160	17	US-10-382-600-7	Sequence 7, Appli
20	2102.5	52.2	6160	17	US-10-382-600-11	Sequence 11, Appli
21	2102.5	52.2	6213	17	US-10-382-600-18	Sequence 18, Appli
22	2102.5	52.2	6213	17	US-10-382-600-21	Sequence 21, Appli
23	2102.5	52.2	6214	17	US-10-382-600-4	Sequence 4, Appli
24	2102.5	52.2	6267	17	US-10-382-600-15	Sequence 15, Appli
25	1654	41.0	1653	9	US-09-792-200B-11	Sequence 11, Appli
26	1387.5	34.4	2373	16	US-10-156-028-1	Sequence 1, Appli
27	1387.5	34.4	2695	14	US-10-028-072-203	Sequence 203, App
28	1387.5	34.4	2695	14	US-10-140-808-203	Sequence 203, App
29	1387.5	34.4	2695	14	US-10-121-049-203	Sequence 203, App
30	1387.5	34.4	2695	14	US-10-123-904-203	Sequence 203, App
31	1387.5	34.4	2695	14	US-10-140-470-203	Sequence 203, App
32	1387.5	34.4	2695	14	US-10-175-746-203	Sequence 203, App
33	1387.5	34.4	2695	14	US-10-176-918-203	Sequence 203, App
34	1387.5	34.4	2695	14	US-10-176-921-203	Sequence 203, App
35	1387.5	34.4	2695	14	US-10-137-865-203	Sequence 203, App
36	1387.5	34.4	2695	14	US-10-140-474-203	Sequence 203, App
37	1387.5	34.4	2695	14	US-10-142-431-203	Sequence 203, App
38	1387.5	34.4	2695	14	US-10-143-114-203	Sequence 203, App
39	1387.5	34.4	2695	14	US-10-142-419-203	Sequence 203, App
40	1387.5	34.4	2695	14	US-10-123-262-203	Sequence 203, App
41	1387.5	34.4	2695	14	US-10-142-423-203	Sequence 203, App
42	1387.5	34.4	2695	14	US-10-121-050-203	Sequence 203, App
43	1387.5	34.4	2695	14	US-10-141-755-203	Sequence 203, App
44	1387.5	34.4	2695	14	US-10-143-032-203	Sequence 203, App
45	1387.5	34.4	2695	14	US-10-123-108-203	Sequence 203, App

#### ALIGNMENTS

##### RESULT 1

US-10-633-202-1  
; Sequence 1, Application US/10633202  
; Publication No. US2004005314A1  
; GENERAL INFORMATION:  
; APPLICANT: CERRETTI, Douglas P.  
; TITLE OF INVENTION: SVPH1-26 DNA AND POLYPEPTIDES  
; FILE REFERENCE: 03260.0036-00304  
; CURRENT APPLICATION NUMBER: US/10/633,202  
; PRIOR FILING DATE: 2003-07-29  
; PRIOR APPLICATION NUMBER: US/09/561,779  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/063,571  
; PRIOR FILING DATE: 1997-10-30  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 1
; LENGTH: 2181
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-633-202-1

Alignment Scores:
  Pred. No.:      0      Length:      2181
  Score:         4030.00  Matches:      726
  Percent Similarity: 100.00%  Conservative: 0
  Best Local Similarity: 100.00%  Mismatches: 0
  Query Match:      100.00%  Indels: 0
  DB:               18      Gaps: 0

US-10-633-202-2 (1-726) x US-10-633-202-1 (1-2181)

Qy 1 MetAlaValGluProLeuValHisIleArgValThrLeuLeuLeuTrpPheGly 20
Db 1 ATGGCAGTGGGTGAGCCCTGGTGACATCAGGGTCACTCTCTGCTGCTCTGGG 60

Qy 21 MetPheLeuSerIleSerGlyHisSerGlnAlaArgProSerGlnTyrPheThrSerPro 40
Db 61 ATGTTTTGTCTATTTCTGGCCACTCTCAGCCAGGCCCTCCAGATATTTCACTTCTCCA 120

Qy 41 GluValValIleProLeuLysValIleSerArgGlyArgGlyAlaLysAlaProGlyTrp 60
Db 121 GAAGTGTGATCCCTTTGAAGGTGATCAGCAGGGGAGAGGTCGAAAGGCTCTGGATGG 180

Qy 61 LeuSerTyrSerLeuArgPheGlyGlyGlnArgTyrIleValHisMetArgValAsnLys 80
Db 181 CTCTCCATATAGCCCTGGGTTGGGGGACAGAGATACATTTCTCACATGAGGGTAAATAAG 240

Qy 81 LeuLeuPheAlaAlaHisLeuProValPheThrTyrThrGluGlnHisAlaLeuLeuGln 100
Db 241 CTGTGTGTGTGTCACACCTTCTGTGTGTTCACCTACACAGAGAGATGCGCTGCCAG 300

Qy 101 AspGlnProPheIleGlnAspAspCysTyrTyrHisGlyTyrValGluGlyValProGlu 120
Db 301 GATCAGCCCTTCATCCAGGATGACTGCTACTACCATGTTATGTGGAGGGGCTCCCTGAG 360

Qy 121 SerLeuValAlaLeuSerThrCysSerGlyGlyPheLeuGlyMetLeuGlnIleAsnAsp 140
Db 361 TCCTTGTGTGGCCCTTAGTACCTGTTCTGGGGGCTTCTTGAAATGCTTACAGATAAATGAC 420

Qy 141 LeuValTyrGluIleLysProIleSerValSerAlaThrPheGluHisLeuValTyrLys 160
Db 421 CTGTGTTATGAATCAAGCAATAGTGTCTTGCCACATTTGAACACCTAGTATATAAG 480

Qy 161 IleAspSerAspThrGlnPheProProMetArgCysGlyLeuThrGluGluLysIle 180
Db 481 ATAGACAGTGATGATACACAGTTTCCACCTATGAGATGTGGGTTAACAGAGAGAAATA 540

Qy 181 AlaHisGlnMetGluLeuGlnLeuSerTyrAsnPheThrLeuLysGlnSerSerPheVal 200
Db 541 GCACACAGATGGAGTTGCAATTTGTCATATAAATTTCACTCTGAAGCAAAAGTTCTTTGTG 600

Qy 201 GlyTrpTrpThrHisGlnArgPheValGluLeuValValValValValValValValVal 220
Db 601 GGCTGGTGGACCCATCAGCGTTTGTGAGCTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 660

Qy 221 LeuPheSerGlnSerAsnAlaThrThrValGlnHisGluValPheAsnValAlaAsnIle 240
Db 661 CTTTTCTCTCAAAGTAATGCAACACACAGTCAGCATCAAGTATTTAAAGTTGTCAATATA 720

Qy 241 ValAspSerPheTyrHisProLeuGluValAspValIleLeuThrGlyIleAspIleTrp 260
Db 721 GTGATTTCTTCTATCATCTCTTGGAGGTTCATGTAATTTTGAAGTGGAAATTTGATATAG 780

Qy 261 ThrAlaSerAsnProLeuProThrSerGlyAspLeuAspAsnValLeuGluAspPheSer 280
Db 781 ACTGCATCAATCCATCTCTTACCAGTGGAGACCTAGATTAATGTTTTAGAGGACTTTTCT 840

Qy 281 IleTrpLysAsnTyrAsnLeuAsnArgLeuGlnHisAspValAlaHisLeuPheIle 300
Db 841 ATTTGGAAGAAATTAACCTTAATAATCGACTACACATGATGTTGCACATCTTTTCATA 900

Qy 301 LysAspThrGlnGlyMetLysLeuGlyValAlaTyrValLysGlyIleCysGlnAsnPro 320
Db 901 AAAGACACACAAGGCATGAAGCTTGGTGTGGCTATGTTAAAGGAATAATGCCAAGATCCT 960

Qy 321 PheAsnThrGlyValAspValPheGluAspAsnArgLeuValValPheAlaIleThrLeu 340
Db 961 TTTAATACCTGGAGTTGATGTTTTTGAAGACACAGGTTGGTGGTTTTTTCGCAATTTCTTG 1020

Qy 341 GlyHisGluLeuGlyHisAsnLeuGlyMetGlnHisAspThrGlnTrpCysValCysGlu 360
Db 1021 GGCCACGAGCTTGGTCTAATTTGGGTATGCAACATGACACCCAGTGGTGTGTGTCGAG 1080

Qy 361 LeuGlnTrpCysIleMetHisAlaTyrArgLysValThrThrLysPheSerAsnCysSer 380
Db 1081 CTACAGTGTGTGATAATGATCCCTATGAAAGGTGACAACTAAATTTTAGCAACTCGAGT 1140

Qy 381 TyrAlaGlnTyrTrpAspSerThrIleSerSerGlyLeuCysIleGlnProProTyr 400
Db 1141 TATGCCAATATTTGGACAGTACTATCAGTAGTGGATTATGATTATTCACCGCTCCATAT 1200

Qy 401 ProGlyAsnIlePheArgLeuLysTyrCysGlyAsnLeuValValGluGlyGluGlu 420
Db 1201 CCAGGGAATATATTTAGACTGAAGTACTGTGGAACTAGTGGTTCGAAGAGGGAGGAA 1260

Qy 421 CysAspCysGlyThrIleArgGlnCysAlaLysAspProCysCysLeuLeuAsnCysThr 440
Db 1261 TGTGACTGTGGAACTACGCGAGTGTGCAAAAGATCCCTGTTGCTGTTAAACTGTACT 1320

Qy 441 LeuHisProGlyValAlaAlaCysAlaPheGlyIleCysCysLysAspCysLysPheLeuPro 460
Db 1321 CTACATCTCTGGGGCTGCTTGTGCTTTTGGAAATATGTTGCAAGACTGCAAAATTTCTGCCA 1380

Qy 461 SerGlyThrLeuCysArgGlnGlnValGlyGluCysAspLeuProGlnTrpCysAsnGly 480
Db 1381 TCAGGAACCTTTATGTAGACAACAAGTTGGTGAATGTGACTTCCAGAGTGGTGCATGG 1440

Qy 481 ThrSerHisGlnCysProAspAspValTyrValGlnAspGlyIleSerCysAsnValAsn 500
Db 1441 ACATCCCATCAATGCCAGATGATGTATGTGAGAGCGGATCTCCTGTGTAATGTGAAT 1500

Qy 501 AlaPheCysTyrGluLysThrCysAsnAsnHisAspIleGlnCysLysGluIlePheGly 520
Db 1501 GCCTTCTGCTATGAAAAGACGTGTAATAACCATGATATACATGTAAGAGATTTTGGC 1560

Qy 521 GlnAspAlaArgSerAlaSerGlnSerCysTyrGlnGluIleAsnThrGlnGlyAsnArg 540
Db 1561 CAAGATGCAAGGAGTGCATCTCAGAGTTGCTACCAAGAAATCAACACCCCAAGGAAACCGT 1620

Qy 541 PheGlyHisCysGlyIleValGlyThrThrTyrValLysCysTrpThrProAspIleMet 560
Db 1621 TTCGGTCACTGGTATTTAGCACACAATATGTAATAATTTTGGACCCCTGTATATCATG 1680

Qy 561 CysGlyArgValGlnCysGluAsnValGlyValIleProAsnLeuIleGluHisSerThr 580
Db 1681 TGTGGAGAGGTTTCAGTGTGAAAATGTGGAGTAATTTCCCAATCTGATAGAGCATTTCTACA 1740

Qy 581 ValGlnGlnPheHisLeuAsnAspThrThrCysTrpGlyThrAspTyrHisLeuGlyMet 600
Db 1741 GTGCAGCAGTTTCCCTCAATGACACCACTTGTGGGGCAGCTGATTTATCAATTTAGGAGATG 1800

Qy 601 AlaIleProAspIleGlyGluValLysAspGlyThrValCysGlyProGluLysIleCys 620
Db 1801 GCTATACCTGATATTCGTGAGGTGAAAGATGGCACAGTATGTGGTCCAGAAAGATCTGC 1860

Qy 621 IleArgLysLysCysAlaSerMetValHisLeuSerGlnAlaCysGlnProLysThrCys 640
Db 1861 ATCCGTAAAGAGTGTCCAGTATGGTTTCATCTGTCAACAGCCTGTGAGCCTTAAGACTGC 1920

Qy 641 AsnMetArgGlyIleCysAsnAsnLysGlnHisCysHisCysAsnHisGluTrpAlaPro 660
Db 1921 ATCCGTAAAGAGTGTCCAGTATGGTTTCATCTGTCAACAGCCTGTGAGCCTTAAGACTGC 1980
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 16, 2005, 00:23:41 ; Search time 432 Seconds  
(without alignments)  
2749.854 Million cell updates/sec

Title: US-10-633-202-2  
Perfect score: 4030  
Sequence: 1 MAVGBPLVHRVTLTLLWFG.....LFLCHLVLFKRTKSKDEEG 726

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool\_p/US10633202/runat\_14062005\_141726\_13120/app.query.fasta\_1.903  
-DB=issued Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10633202@cgn 1.105 @runat\_14062005\_141726\_13120 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C	1	4030	100.0	35837	4 US-09-949-016-15232
	2	2435	60.4	2169	4 US-09-617-145-1
	3	2435	60.4	3557	4 US-09-949-016-547
	4	2376	59.0	7557	4 US-09-949-016-12289
	5	2108.5	52.3	2959	4 US-09-949-016-3341
	6	2108.5	52.3	2959	4 US-09-949-016-3342
	7	2108.5	52.3	2959	4 US-09-949-016-3343
	8	2108.5	52.3	50850	4 US-09-949-016-15083
	9	2108.5	52.3	50850	4 US-09-949-016-15084
	10	2108.5	52.3	50850	4 US-09-949-016-15085
	11	1387.5	34.4	2373	4 US-09-949-016-2298
	12	1387.5	34.4	2373	4 US-09-949-016-2299

13	1387.5	34.4	6373	4	US-09-949-016-12830	Sequence 12830, A
14	1387.5	34.4	6373	4	US-09-949-016-12888	Sequence 12888, A
15	1387.5	34.4	6374	4	US-09-949-016-14040	Sequence 14040, A
16	1387.5	34.4	6374	4	US-09-949-016-14041	Sequence 14041, A
17	1386.5	34.4	2346	4	US-09-949-016-1088	Sequence 1088, Ap
18	1385.5	34.4	2373	4	US-09-949-016-1146	Sequence 1146, Ap
19	1317	32.7	2207	4	US-09-949-016-4593	Sequence 4593, Ap
20	1175.5	29.2	3758	4	US-09-976-594-491	Sequence 491, App
21	1009.5	25.0	5062	4	US-09-949-016-499	Sequence 499, App
22	980.5	24.3	4471	4	US-09-949-016-4910	Sequence 4910, Ap
23	955.5	23.7	2976	4	US-09-949-016-5358	Sequence 5358, Ap
24	955.5	23.7	3261	4	US-09-949-016-2391	Sequence 2391, Ap
25	924	22.9	3468	3	US-09-632-098-3	Sequence 3, Appli
26	924	22.9	3468	4	US-10-177-308-3	Sequence 3, Appli
27	917.5	22.8	3582	4	US-09-548-797B-3	Sequence 3, Appli
28	911	22.6	2650	2	PCT-US95-07295-7	Sequence 7, Appli
29	911	22.6	2650	5	PCT-US95-07295-7	Sequence 7, Appli
30	885	22.0	3390	4	US-09-548-797B-2	Sequence 2, Appli
31	884.5	21.9	2556	4	US-10-000-489-69	Sequence 69, Appl
32	882.5	21.9	2325	4	US-09-786-256C-16	Sequence 16, Appl
33	882.5	21.9	2828	4	US-09-786-256C-31	Sequence 31, Appl
34	878	21.8	2373	1	US-08-264-101-1	Sequence 1, Appli
35	878	21.8	2373	2	US-08-765-243-1	Sequence 1, Appli
36	878	21.8	2373	5	PCT-US95-07295-1	Sequence 5, Appli
37	876.5	21.7	2553	2	US-08-765-243-5	Sequence 5, Appli
38	876.5	21.7	2553	3	PCT-US95-07295-5	Sequence 5, Appli
39	855.5	21.2	3431	3	US-09-632-098-1	Sequence 1, Appli
40	855.5	21.2	3431	4	US-10-177-308-1	Sequence 1, Appli
41	839	20.8	2648	2	US-08-836-443-1	Sequence 1, Appli
42	826	20.5	3994	4	US-09-738-945-7	Sequence 7, Appli
43	815.5	20.2	3271	4	US-09-548-797B-1	Sequence 1, Appli
44	808	20.0	1851	3	US-09-608-790-2	Sequence 2, Appli
45	799.5	19.8	2251	2	US-08-836-443-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-949-016-15232/c  
; Sequence 15232, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 15232  
; LENGTH: 35837  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15232

Alignment Scores:  
Pred. No.: 0 Length: 35837  
Score: 4030.00 Matches: 726  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-633-202-2 (1-726) x US-09-949-016-15232 (1-35837)

QY 1 MetAlaValGlyGluProLeuValHisIleArgValThrLeuLeuLeuTrpPheGly 20

Db 2336 ATGCGAGTGGGTGAGCCCTGGTGCACATCAGGGTCACTCTCTGCTGCTCTGCTTGGG 2277  
Qy 21 MetPheLeuSerIleSerGlyHisSerGlnAlaArgProSerGlnTyrPheThrSerPro 40  
Db 2276 ATGTTTTTGTCTATTTCTGGCCACTCTCAGGCCAGGCCCTCCCAAGTATTTCACTTCTCCA 2217  
Qy 41 GluValValIleProLeuLysValIleSerArgGlyArgGlyAlaLysAlaProGlyTyr 60  
Db 2216 GAAGTGGTGATCCCTTTGAAGGTGATCAGCAGGGGAGAGGTGCANAGGCTCTCGNATGG 2157  
Qy 61 LeuSerTyrSerLeuArgPheGlyGlyGlnArgTyrIleValHisMetArgValAsnLys 80  
Db 2156 CTCTCCATATAGCCCTGGGTCAGAGGAGAGATACATTTGTCCACATGAGGGTAAATAAG 2097  
Qy 81 LeuLeuPheAlaAlaHisLeuProValPheThrTyrThrGluGlnHisAlaLeuLeuGln 100  
Db 2096 CTGTTGTTTGTGTCACACCTTCTCTGTGTTCACCTACACAGAGCAGATGCCCTGCTCCAG 2037  
Qy 101 AspGlnProPheIleGlnAspAspCysTyrTyrHisGlyTyrValGluGlyValProGlu 120  
Db 2036 GATCAGCCCTTCATCCAGGATGACTGCTACTACCATGGTTATGTGGAGGGGTCCTCGAG 1977  
Qy 121 SerLeuValAlaLeuSerThrCysSerGlyGlyPheLeuGlyMetLeuGlnIleAsnAsp 140  
Db 1976 TCCCTGGTGGCTTAGTACCTGTTCTGGGGCTTCTTGGAAATGCTACAGATAAATGAC 1917  
Qy 141 LeuValTyrGluIleLysProIleSerValSerAlaThrPheGluHisLeuValTyrLys 160  
Db 1916 CTTGTTTATGAAATCAAGCCCAATTAGTGTTCGCCACATTTGAAACCTAGTATATAAG 1857  
Qy 161 IleAspSerAspThrGlnPheProProMetArgCysGlyLeuThrGluLysIle 180  
Db 1856 ATACACAGTGNATGATACAGTTTCCACCTATGAGATGTGGTTAACAGAGAGAAATA 1797  
Qy 181 AlaHisGlnMetGluLeuGlnLeuSerTyrAsnPheThrLeuLysGlnSerSerPheVal 200  
Db 1796 GCACACAGATGGAGTTGCAATTTGTCATATAATTTCACTCTGAAGCAAAAGTTCTTTGTG 1737  
Qy 201 GlyTyrTrpThrHisGlnArgPheValGluLeuValValValValAspAsnIleArgTyr 220  
Db 1736 GGCTGGTGGACCCATCAGCGGTTGTTGAGCTGGTAGTGGTGGGATAATATATAGATAT 1677  
Qy 221 LeuPheSerGlnSerAsnAlaThrThrValGlnHisGluValPheAsnValAlaAsnIle 240  
Db 1676 CTTTTCTCBAAGTAATGCACACACAGTGCAGATGAGTATTTAACTGTTGTCAATATA 1617  
Qy 241 ValAspSerPheTyrHisProLeuGluValAspValIleLeuThrGlyIleAspIleTyr 260  
Db 1616 GTGGATTCTCTATCATCTCTTGGAGGTTGATGTAATTTTGAAGTGAATGATATATGG 1557  
Qy 261 ThrAlaSerAsnProLeuProThrSerGlyAspLeuAspAsnValLeuGluAspPheSer 280  
Db 1556 ACTGCATCAAAATCCACTTCTTACCAGTGGAGACCTAGATAATGTTTTAGAGGACTTTTCT 1497  
Qy 281 IleTrpLysAsnTyrAsnLeuAsnAsnArgLeuGlnHisAspValAlaHisLeuPheIle 300  
Db 1496 ATTTGGGAAGATTAATACCTTAATATCGACTACACATGATGTTGCACATCTTTTCATA 1437  
Qy 301 LysAspThrGlnGlyMetLysLeuGlyValAlaTyrValLysGlyIleCysGlnAsnPro 320  
Db 1436 AAAGACACACAAGGCATGAGCTGGTGTGGCTTATGTTAAAGGAATATGCCAGATCCT 1377  
Qy 321 PheAsnThrGlyValAspValPheGluAspAsnArgLeuValValPheAlaIleThrLeu 340  
Db 1376 TTTAAATAGTGGATTGATGTTTTTGAAGACACACAGGTGGTGGCTTTTGAATTTACTTTG 1317  
Qy 341 GlyHisGluLeuGlyHisAsnLeuGlyMetGlnHisAspThrGlnTyrCysValCysGlu 360  
Db 1316 GGCCACAGGCTTGGTCATATTTGGGTATGCACATGACACCCAGTGGTGTGTGTGGAG 1257  
Qy 361 LeuGlnTrpCysIleMetHisAlaTyrArgLysValThrThrLysPheSerAsnCysSer 380

Db 1256 CTACAGTGGTGCATATATGCATGCTATAGAAAGGTGACAACATAAATTTAGCAACTCAGT 1197  
Qy 381 TyrAlaGlnTyrTrpAspSerThrIleSerSerGlyLeuCysIleGlnProProProTyr 400  
Db 1196 TATGCCCAATATTTGGGACAGTACTATCAGTAGTGGATTAATGATTATCAACCGCTCCATAT 1137  
Qy 401 ProGlyAsnIlePheArgLeuLysTyrCysGlyAsnLeuValValGluGluGlyGlu 420  
Db 1136 CCAGGGAATATATTTAGACTGAAGTACTGTGGGAATCTAGTGGTTGAAGAGGGAGGAA 1077  
Qy 421 CysAspCysGlyThrIleArgGlnCysAlaLysAspProCysCysLeuLeuAsnCysThr 440  
Db 1076 TGTGACTGTGGAACCACTACGCGAGTGTGCAAAAGATCCCTGTTGTCTGTTAAACTGTACT 1017  
Qy 441 LeuHisProGlyValAlaAlaCysAlaPheGlyIleCysCysLysAspCysLysPheLeuPro 460  
Db 1016 CTACATCTCGGGCTCTTGTGCTTTTGGAAATATGTTGAAGACTGCAAAATTTCTGCCA 957  
Qy 461 SerGlyThrLeuCysArgGlnGlnValGlyGluCysAspLeuProGluTrpCysAsnGly 480  
Db 956 TCAGGAACCTTATGTAGACACAACAAAGTTGGTGAATGTGACCTTCCAGAGTGGTGAATGG 897  
Qy 481 ThrSerHisGlnCysProAspAspValTyrValGlnAspGlyIleSerCysAsnValAsn 500  
Db 896 ACATCCCATCAATGCCCAGATGATGTATGTGTCAGAGCGGATCTCTCTGTAATGTGAAT 837  
Qy 501 AlaPheCysTyrGluLysThrCysAsnAsnHisAspIleGlnCysLysGluIlePheGly 520  
Db 836 GCCTTCTGCTATGAAAAGACGTGTAATAACCATGATATACATGTAAGAGATTTTGGC 777  
Qy 521 GlnAspAlaArgSerAlaSerGlnSerCysTyrGlnGluIleAsnThrGlnGlyAsnArg 540  
Db 776 CRAAGTGAAGGAGTGCATCTCAGAGTTGCTACCAAGAAATCAACACCCCAAGGAAACCGT 717  
Qy 541 PheGlyHisCysGlyIleValGlyThrTyrValLysCysTrpThrProAspIleMet 560  
Db 716 TTCCGTCACCTGGTATTTAGGCACACATATGTAATAATGTTGGACCCCTGTATATCATG 657  
Qy 561 CysGlyArgValGlnCysGluAsnValGlyValIleProAsnLeuIleGluHisSerThr 580  
Db 656 TGTGGAGGAGGTTCAAGTGTGAAAATGTTGGAGTAATTTCCCAATCTGATAGACATTTCTACA 597  
Qy 581 ValGlnGlnPheHisLeuAsnAspThrThrCysTyrGlyThrAspTyrHisLeuGlyMet 600  
Db 596 GTGCACAGTTTCCACCTCAATGACACCACTTGTGCGGGCACTGATTTATCATTTAGGGATG 537  
Qy 601 AlaIleProAspIleGlyGluValLysAspGlyThrValCysGlyProGluLysIleCys 620  
Db 536 GCTATACCTGATATTTGGTGGAGGTGAAAGATGGCACAGTATGTGGTCCAGAAAAGATCTGC 477  
Qy 621 IleArgLysLysCysAlaSerMetValHisLeuSerGlnAlaCysGlnProLysThrCys 640  
Db 476 ATCCGTAAAGAGTGTCCAGTATGGTTTCATCTGTCAAGCCCTGTGAGCTTAAGACCTGC 417  
Qy 641 AsnMetArgGlyIleCysAsnAsnLysGlnHisCysHisCysAsnHisGluTrpAlaPro 660  
Db 416 AACATGAGGGGAATCTGCACACCAACAACTGTCTGCAACCATGAAATGGGCACCC 357  
Qy 661 ProTyrCysLysAspLysGlyTyrGlyLysArgAlaAspSerGlyProProProLysAsn 680  
Db 356 CCATACCTGCAAGGACAAAGGCTATGGAGGTAGTGTCTGATAGTGGCCACCTCTCTAAGAAC 297  
Qy 681 AsnMetGluGlyLeuAsnValMetGlyLysLeuArgTyrLeuSerLeuLeuCysLeuLeu 700  
Db 296 AACATGGAAGGATTAATGTGATGGGAAAGTTGCGTTTACTCTGTCACTATTTGCGCTTCT 237  
Qy 701 ProLeuValAlaPheLeuLeuPheCysLeuHisValLeuPheLysLysArgThrLysSer 720  
Db 236 CCTTTGTTGCTTTTATTTATTTTGTCTTACATGCTCTTTTAAAGAAACGACACAAAAGT 177  
Qy 721 LysGluAspGluGly 726  
Db 176 AAAGAAGATGAAGAAGGA 159



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 20:14:51 ; Search time 110 Seconds  
(without alignments)  
2530.007 Million cell updates/sec

Title: US-10-633-202-2  
Perfect score: 4030  
Sequence: 1 MAVGPELVHVRVTLMLFPG.....LFLHLVFKRTKSKDEEG 726

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1710399 seqs, 383334425 residues

Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10F\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4030	100.0	726	15 US-10-633-202-2	Sequence 2, Appli
2	4030	100.0	776	17 US-09-978-418-6	Sequence 6, Appli
3	4030	100.0	776	17 US-10-485-231-6	Sequence 6, Appli
4	4004	99.4	726	14 US-10-156-028-3	Sequence 3, Appli
5	4004	99.4	726	17 US-10-491-997-106	Sequence 106, App
6	2435	60.4	722	14 US-10-265-125-2	Sequence 2, Appli
7	2435	60.4	722	15 US-10-074-978A-289	Sequence 289, App
8	2427	60.2	722	15 US-10-074-978A-288	Sequence 288, App
9	2102.5	52.2	766	15 US-10-664-456-12	Sequence 12, Appl
10	2102.5	52.2	787	15 US-10-664-456-13	Sequence 13, Appl
11	2102.5	52.2	820	15 US-10-664-456-14	Sequence 14, Appl

12	2102.5	52.2	820	16	US-10-408-765A-1252	Sequence 1252, Ap
13	2050	50.9	729	15	US-10-074-978A-287	Sequence 287, App
14	1654	41.0	446	9	US-09-792-2008-12	Sequence 12, Appl
15	1646	40.8	535	15	US-10-633-202-3	Sequence 3, Appli
16	1387.5	34.4	790	14	US-10-028-072-204	Sequence 204, App
17	1387.5	34.4	790	14	US-10-140-808-204	Sequence 204, App
18	1387.5	34.4	790	14	US-10-121-049-204	Sequence 204, App
19	1387.5	34.4	790	14	US-10-123-904-204	Sequence 204, App
20	1387.5	34.4	790	14	US-10-140-470-204	Sequence 204, App
21	1387.5	34.4	790	14	US-10-175-746-204	Sequence 204, App
22	1387.5	34.4	790	14	US-10-176-918-204	Sequence 204, App
23	1387.5	34.4	790	14	US-10-176-921-204	Sequence 204, App
24	1387.5	34.4	790	14	US-10-137-855-204	Sequence 204, App
25	1387.5	34.4	790	14	US-10-140-474-204	Sequence 204, App
26	1387.5	34.4	790	14	US-10-142-431-204	Sequence 204, App
27	1387.5	34.4	790	14	US-10-143-114-204	Sequence 204, App
28	1387.5	34.4	790	14	US-10-142-419-204	Sequence 204, App
29	1387.5	34.4	790	14	US-10-123-262-204	Sequence 204, App
30	1387.5	34.4	790	14	US-10-142-423-204	Sequence 204, App
31	1387.5	34.4	790	14	US-10-121-050-204	Sequence 204, App
32	1387.5	34.4	790	14	US-10-141-755-204	Sequence 204, App
33	1387.5	34.4	790	14	US-10-143-032-204	Sequence 204, App
34	1387.5	34.4	790	14	US-10-123-108-204	Sequence 204, App
35	1387.5	34.4	790	14	US-10-123-236-204	Sequence 204, App
36	1387.5	34.4	790	14	US-10-123-261-204	Sequence 204, App
37	1387.5	34.4	790	14	US-10-140-921-204	Sequence 204, App
38	1387.5	34.4	790	14	US-10-140-928-204	Sequence 204, App
39	1387.5	34.4	790	14	US-10-121-045-204	Sequence 204, App
40	1387.5	34.4	790	14	US-10-123-292-204	Sequence 204, App
41	1387.5	34.4	790	14	US-10-123-903-204	Sequence 204, App
42	1387.5	34.4	790	14	US-10-124-819-204	Sequence 204, App
43	1387.5	34.4	790	14	US-10-124-822-204	Sequence 204, App
44	1387.5	34.4	790	14	US-10-140-925-204	Sequence 204, App
45	1387.5	34.4	790	14	US-10-160-498-204	Sequence 204, App

ALIGNMENTS

RESULT 1  
US-10-633-202-2  
; Sequence 2, Application US/10633202  
; Publication NO. US20040053141A1  
; GENERAL INFORMATION:  
; APPLICANT: CERRETTI, Douglas P.  
; TITLE OF INVENTION: SVPH1-26 DNA AND POLYPEPTIDES  
; FILE REFERENCE: 03260.0036-00304  
; CURRENT FILING DATE: 2003-07-29  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US/09/561,779  
; PRIOR FILING DATE: 1997-10-30  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-633-202-2

Query Match	100.0%	Score 4030	DB 15	Length 726
Best Local Similarity	100.0%	Pred. No. 0		
Matches 726	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	MAVGPELVHVRVTLMLFPGMFLSISGHSQARPSQYFTSPVPLKVISRGKAKPGW	60	
Db	1	MAVGPELVHVRVTLMLFPGMFLSISGHSQARPSQYFTSPVPLKVISRGKAKPGW	60	
Qy	61	LSVSLRFGGQRIYVHVRVNLKFAAHLPVFTTTEQHALLQDQPFIODDCYHYHVGVEPE	120	
Db	61	LSVSLRFGGQRIYVHVRVNLKFAAHLPVFTTTEQHALLQDQPFIODDCYHYHVGVEPE	120	

```
Qy 121 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 180
Db 121 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 180
Qy 181 AHQMEIQLSNTFLKQSSFGVWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFNVNI 240
Db 181 AHQMEIQLSNTFLKQSSFGVWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFNVNI 240
Qy 241 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 300
Db 241 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 300
Qy 301 KDTQGMKLGVAAYKGIQNPFTGVDVFEDNRLVFAITLGHGHLNGLQHDQTQMCVCE 360
Db 301 KDTQGMKLGVAAYKGIQNPFTGVDVFEDNRLVFAITLGHGHLNGLQHDQTQMCVCE 360
Qy 361 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 420
Db 361 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 420
Qy 421 CDGCTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 480
Db 421 CDGCTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 480
Qy 481 TSHQCDDVYVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQIBINTQGNR 540
Db 481 TSHQCDDVYVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQIBINTQGNR 540
Qy 541 FGHCGIVGTYVKCWTDPDIMGCRVQCNVGI PNLIHSTVQQFHLNDTTCWGTDYHLGM 600
Db 541 FGHCGIVGTYVKCWTDPDIMGCRVQCNVGI PNLIHSTVQQFHLNDTTCWGTDYHLGM 600
Qy 601 AIPDGEVKDGTVCGEPEKICIRKKCASVHLSOACQPKTCNMREGICNNKQHCNHEWAP 660
Db 601 AIPDGEVKDGTVCGEPEKICIRKKCASVHLSOACQPKTCNMREGICNNKQHCNHEWAP 660
Qy 661 PYCKDKYGGSGADSGPPPKNNMEGLNVMGKRLYLSLLCLLPLVAFLLCFLHVLFKKRTKS 720
Db 661 PYCKDKYGGSGADSGPPPKNNMEGLNVMGKRLYLSLLCLLPLVAFLLCFLHVLFKKRTKS 720
Qy 721 KEDEEG 726
Db 721 KEDEEG 726

RESULT 2
US-09-978-418-6
; Sequence 6, Application US/09978418
; Publication No. US20030118997A1
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephan
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 142.US5.REG
; CURRENT APPLICATION NUMBER: US/09/978,418
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/311,305
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/314,734
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/318,204
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/326,470
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: JPatent
; SEQ ID NO 6
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-418-6
Query Match 100.0%; Score 4030; DB 10; Length 776;
```

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAVGEPLVHVRVTLTLLWFGMELISGHSQARPSQYFTSPVVIPLKVISRGRGAKAPGW 60
Db 51 MAVGEPLVHVRVTLTLLWFGMELISGHSQARPSQYFTSPVVIPLKVISRGRGAKAPGW 110
Qy 61 LYSLSRFGGORYIVHMRVNKLLFAAHLPVFTYTEQHALLQDQDQFIQDDCYHYHGYVEGVE 120
Db 111 LYSLSRFGGORYIVHMRVNKLLFAAHLPVFTYTEQHALLQDQDQFIQDDCYHYHGYVEGVE 170
Qy 121 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 180
Db 171 SLVALSTCSGGFLGMLQINDLVYKIPISVSATFEHLVYKIDSDDTQFPMPRCGLTEEKI 230
Qy 181 AHQMEIQLSNTFLKQSSFGVWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFNVNI 240
Db 231 AHQMEIQLSNTFLKQSSFGVWTHQRFVELVVDNIRYLFQSQSNATTVOHEVFNVNI 290
Qy 241 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 300
Db 291 VDSFYHPLVDVILTGIDIWTSNPLPTSGDLNVLDFSIWKYNNLNRLQHDVAHLFI 350
Qy 301 KDTQGMKLGVAAYKGIQNPFTGVDVFEDNRLVFAITLGHGHLNGLQHDQTQMCVCE 360
Db 351 KDTQGMKLGVAAYKGIQNPFTGVDVFEDNRLVFAITLGHGHLNGLQHDQTQMCVCE 410
Qy 361 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 420
Db 411 LQWCIHAYRKVTKFSNCSYAQYWDSTISSGLCIQPPYPGNI FRLKYCGNLVVEEGE 470
Qy 421 CDGCTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 480
Db 471 CDGCTIRQCAKDPCCLLNCTLHPGAACAFGICCKDCKFLPSGTLCRQQVGECDLPEWCNG 530
Qy 481 TSHQCDDVYVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQIBINTQGNR 540
Db 531 TSHQCDDVYVQDGISCNVNAFCYEKTCNNHDIQCKEIFGQDARSASQSCYQIBINTQGNR 590
Qy 541 FGHCGIVGTYVKCWTDPDIMGCRVQCNVGI PNLIHSTVQQFHLNDTTCWGTDYHLGM 600
Db 591 FGHCGIVGTYVKCWTDPDIMGCRVQCNVGI PNLIHSTVQQFHLNDTTCWGTDYHLGM 650
Qy 601 AIPDGEVKDGTVCGEPEKICIRKKCASVHLSOACQPKTCNMREGICNNKQHCNHEWAP 660
Db 651 AIPDGEVKDGTVCGEPEKICIRKKCASVHLSOACQPKTCNMREGICNNKQHCNHEWAP 710
Qy 661 PYCKDKYGGSGADSGPPPKNNMEGLNVMGKRLYLSLLCLLPLVAFLLCFLHVLFKKRTKS 720
Db 711 PYCKDKYGGSGADSGPPPKNNMEGLNVMGKRLYLSLLCLLPLVAFLLCFLHVLFKKRTKS 770
Qy 721 KEDEEG 726
Db 771 KEDEEG 776

RESULT 3
US-10-485-231-6
; Sequence 6, Application US/10485231
; Publication No. US20050119171A1
; GENERAL INFORMATION:
; APPLICANT: GENSET
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 142.WO1
; CURRENT APPLICATION NUMBER: US/10/485,231
; CURRENT FILING DATE: 2004-01-29
; PRIOR APPLICATION NUMBER: 60/311,305
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/314,734
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/318,204
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/326,470
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 19:52:20 ; Search time 56 Seconds  
(without alignments)  
967.772 Million cell updates/sec

Title: US-10-633-202-2  
Perfect score: 4030  
Sequence: 1 MAVGSLVHVRVTLMLWFG.....LFCLHVLPKKTKSKEDDEG 726

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/1/aaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/prodata/1/aaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/prodata/1/aaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/prodata/1/aaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/prodata/1/aaa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/prodata/1/aaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2435	60.4	722	US-09-617-145-2	Sequence 2, Appli
2	2435	60.4	722	US-09-949-016-6418	Sequence 6418, Ap
3	2108.5	52.3	826	US-09-949-016-9212	Sequence 9212, Ap
4	2108.5	52.3	826	US-09-949-016-9213	Sequence 9213, Ap
5	2108.5	52.3	826	US-09-949-016-9214	Sequence 9214, Ap
6	1387.5	34.4	790	US-09-949-016-8169	Sequence 8169, Ap
7	1387.5	34.4	790	US-09-949-016-8170	Sequence 8170, Ap
8	1386.5	34.4	781	US-09-949-016-6959	Sequence 6959, Ap
9	1385.5	34.4	790	US-09-949-016-7017	Sequence 7017, Ap
10	1314.5	32.6	705	US-09-949-016-10464	Sequence 10464, A
11	1009.5	25.0	909	US-09-949-016-6370	Sequence 6370, Ap
12	955.5	23.7	953	US-09-949-016-8262	Sequence 8262, Ap
13	955.5	23.7	968	US-09-949-016-11229	Sequence 11229, A
14	924	22.9	812	US-09-632-098-4	Sequence 4, Appli
15	924	22.9	812	US-10-177-308-4	Sequence 4, Appli
16	917.5	22.8	849	US-09-548-797B-6	Sequence 6, Appli
17	910	22.6	734	US-08-765-243-8	Sequence 8, Appli
18	910	22.6	734	PCT-US95-07295-8	Sequence 8, Appli
19	885	22.0	787	US-09-348-797B-5	Sequence 5, Appli
20	882.5	21.9	775	US-09-786-256C-15	Sequence 15, Appl
21	882.5	21.9	775	US-09-786-256C-32	Sequence 32, Appl
22	879.5	21.8	787	US-10-000-489-70	Sequence 70, Appl
23	878	21.8	651	US-08-264-101-2	Sequence 2, Appli
24	878	21.8	651	US-08-765-243-2	Sequence 2, Appli
25	878	21.8	651	PCT-US95-07295-2	Sequence 2, Appli
26	873.5	21.7	735	US-08-765-243-6	Sequence 6, Appli
27	873.5	21.7	735	PCT-US95-07295-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-617-145-2  
; Sequence 2, Application US/09617145  
; Patent No. 6485956  
; GENERAL INFORMATION:  
; APPLICANT: Cerretti, Douglas P.  
; TITLE OF INVENTION: SVPHI-8 DNA and Polypeptides  
; FILE REFERENCE: 03260.0050-00304  
; CURRENT APPLICATION NUMBER: US/09/617,145  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/071,505  
; PRIOR FILING DATE: 1998-01-14  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 722  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-617-145-2

Query Match 60.4%; Score 2435; DB 4; Length 722;

Best Local Similarity 61.6%; Pred. No. 2e-207;

Matches 444; Conservative 101; Mismatches 172; Indels 4; Gaps 4;

Qy	1	MAVGSLVHVRVTLMLWFGFLSISGHSQAPSPQYFTSPVIVPLKVISRGKAPGM	60
Db	1	MAVDGLVYIRVTLMLLWGLVFLSISGYSQAPSPQYFTSPVIVPLKVISRGKAPGM	60
Qy	61	LSYSLRFGGQRYIVHMYNKLFAAHLPVFTVTEQHALLQDPFIQDDCYHYGYEGVPE	120
Db	61	LSYSLRFGGQRYIVHMYNKLFAAHLPVFTVTEQHALLQDPFIQDDCYHYGYEGVPE	120
Qy	121	SLVALSTCGGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDTQPPMRCGLTEEKI	180
Db	121	SLVVSACFGGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDTQPPMRCGLTEEKI	180
Qy	181	A-HQMLQSLNFTYTKQSSFCVGNWTHQRPVELVVDVVDVVDVVDVVDVVDVVDVVDV	239
Db	181	ARQQLFEFEAEASALEPKSAGDWTHAWFLELVVDVVDVVDVVDVVDVVDVVDVVDV	240
Qy	240	IVDSFVHPLVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDV	299
Db	241	IVDSVYQQLGTVIILIGIEIMNQGVFPMVTSIEQVLDNFQWQKQISL-SQLOHAAHNF	298
Qy	300	IKDTCQMKLGVAIVKGIQNPENTGVDFEDNRLVVFATLGHGLHNLGMQHDTCWCYC	359
Db	299	IKNSLISILGLAYVAGICRPDIDCVDFNFGQDTWSLFWANTVAHGLHTLGMQHDDEFCFC	358
Qy	360	ELQWCMIMHAYRKVTKFNCSVAQVWDSTISSGLCIQPPYPGNTFRLKYCNGLVVEEGE	419

```
Db 359 GERCIMNTFRVPAEKFTNCSYADFMKTTTLNQSGSLNPPRLGEIFMLKRCGNGVVEREE 418
Qy 420 EDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGTLCRQQVGECDLPEWCN 479
Db 419 QDCGVSQOCEODACCLNCTLRGAACAFGLCKCKFMPSGELCHQEVNECDLPEWCN 478
Qy 480 GTSHQCPDDVYVQDGISCNVNAFCYKTCNNHDIQCKEIFGQDARSASQSCYQEIINTQGN 539
Db 479 GTSHQCPEDRYVQDGIPCSDSAYCYQKRCNNHDOHCREIFGDKAKSASQNCYKEINSQGN 538
Qy 540 RFHCGIGVGTYYVKCWPDMIMCGRVQCEVGVTPNLLIEHSTVQOQFHLNDTTCWGTDYHLG 599
Db 539 RFHCGINGITYLKCHISDVFCGRVQCEVNRDIPLLQDHFTLQHTHINGVTCWGDYHLR 598
Qy 600 MAIPDIEGVKDGTVCGPEKICIRKCKASVMHLSQACQPKTCNMRGICNNKHQCHCHNEWA 659
Db 599 MNISDIEGVKDGTVCGPGKICIHKKCVSLVSLSHVCLPETCNMKGICNNKHCHCHCGYWS 658
Qy 660 PPYCKDKYGGSDSGPPPKNNMEGLNMGKRLYLSLLCLPLVAFLLFCLHVLFKKRTK 719
Db 659 PPYQHRGYGSGSDSGPASAKRGVFLPLI-VIPSLSVLTFTVGLMYLRLQCSGPKETK 717
Qy 720 S 720
Db 718 A 718

RESULT 2
US-09-949-016-6418
; Sequence 6418, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6418
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6418

Query Match 60.4%; Score 2435; DB 4; Length 722;
Best Local Similarity 61.6%; Pred. No. 2e-207;
Matches 444; Conservative 101; Mismatches 172; Indels 4; Gaps 4;

Qy 1 MAYGEPIVHVRVILLWFMFLSISGHSQARPSQFTSPSEVVIPLKVISRGKAPGW 60
Db 1 MAVDGLTVIRVILLWLVFLSISGYCQAGFSQHTSPSEVVIPLKVISRGKAPGW 60
Qy 61 LSYSLRFGGQRYIVHMRVKNLLFAAHLPVFTYTEQHALLODQPFIODDCYHYGVGEVPE 120
Db 61 LSYSLRFGGQRYVHMRVKNLLFAAHLPVFTYTEQHALLODQPFIODDCYHYGVGEVPE 120
Qy 121 SLVALSCSGGFGMLQINDLVVEIKPISVSATFEHLVYKIDSDDTQFPFMRGLTEEKI 180
Db 121 SLVVFSAACGFGGVLKISGLTVEIEPIRSATFEHLVYKVNNSNETQFPAMRGLTEKEV 180
Qy 181 A-HQMEQLQLSYNFTLKQSSFVGWTHQRFVELVVDNIRYLSQSNATTVQHEVFNVN 239
Db 181 ARQOLEFEAEENALEPKSAGDWTWAFLELVVDNIRYLSQSNATTVQHEVFNVN 240
Qy 240 IVDSFYHPLEVDVILTGIDITWASNPLPTSGDLONVLEDFSIWKNYNLNNRLQDVAHLF 299
```

```
Db 241 IVDSMTYQQLGTYYILIGIBIWNQGNVFPMT-SIEQVLNDFSQWKQISL-SQLQDAHMF 298
Qy 300 IKDTQGMKLGVAVKVIGICQNPENTGVDFEDNRLVVFATLTGHELGHNLGMQHDTCWCVC 359
Db 299 IKNSLSIILGLAYVAGICRPPIDCGVDNFGQDWTSLFANTVAHELGHTLGMQHDEEFCC 358
Qy 360 ELQWCIIMHAYRKVTTKFSNCSYQYWDSTISSGLCTIQPPYPGNIIFRLKYCGNLVVEEGE 419
Db 359 GERGCIMNTFRVPAEKFTNCSYADFMKTTTLNQSGSLNPPRLGEIFMLKRCGNGVVEREE 418
Qy 420 EDCGTTIRQCAKDPCCLLNCTLHPGAACAFGICCKCKFLPSGTLCRQQVGECDLPEWCN 479
Db 419 QDCGVSQOCEODACCLNCTLRGAACAFGLCKCKFMPSGELCHQEVNECDLPEWCN 478
Qy 480 GTSHQCPDDVYVQDGISCNVNAFCYKTCNNHDIQCKEIFGQDARSASQSCYQEIINTQGN 539
Db 479 GTSHQCPEDRYVQDGIPCSDSAYCYQKRCNNHDOHCREIFGDKAKSASQNCYKEINSQGN 538
Qy 540 RFHCGIGVGTYYVKCWPDMIMCGRVQCEVGVTPNLLIEHSTVQOQFHLNDTTCWGTDYHLG 599
Db 539 RFHCGINGITYLKCHISDVFCGRVQCEVNRDIPLLQDHFTLQHTHINGVTCWGDYHLR 598
Qy 600 MAIPDIEGVKDGTVCGPEKICIRKCKASVMHLSQACQPKTCNMRGICNNKHQCHCHNEWA 659
Db 599 MNISDIEGVKDGTVCGPGKICIHKKCVSLVSLSHVCLPETCNMKGICNNKHCHCHCGYWS 658
Qy 660 PPYCKDKYGGSDSGPPPKNNMEGLNMGKRLYLSLLCLPLVAFLLFCLHVLFKKRTK 719
Db 659 PPYQHRGYGSGSDSGPASAKRGVFLPLI-VIPSLSVLTFTVGLMYLRLQCSGPKETK 717
Qy 720 S 720
Db 718 A 718

RESULT 3
US-09-949-016-9212
; Sequence 9212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9212
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9212

Query Match 52.3%; Score 2108.5; DB 4; Length 826;
Best Local Similarity 51.8%; Pred. No. 2.5e-178;
Matches 381; Conservative 112; Mismatches 217; Indels 25; Gaps 6;

Qy 5 EPLVHVRVILLWFMFLSISGHSQARPSQFTSPSEVVIPLKVISRGKAPGWLSYS 64
Db 2 EAFNKKMLLLHCLGFLVLSGSHIQDEHPQYHSPDPVIVPVRITGTTTGTMTPTPGWLSYI 61
Qy 65 LRFGGORYVHMRVKNLLFAAHLPVFTYTEQHALLODQPFIODDCYHYGVGEVPESLVA 124
Db 62 LPFGGQKHIIHKVKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYHYGVGEVPEPSLVS 121
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 16:25:28 ; Search time 123 Seconds  
(without alignments)  
2282.830 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 MAVGRLVIRVTLMLWF.....LFLHLVLFKRTKSEDEEG 726

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003s:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4030	100.0	726	2	AAY17413 Human SVP
2	4030	100.0	776	6	ABR39425 Human GEN
3	4004	99.4	726	2	AAY03223 Amino aci
4	4004	99.4	726	7	ADC78877 Human SVP
5	2435	60.4	722	2	AAY28655 Human SVP
6	2435	60.4	722	6	ABG76200 Human sna
7	2102.5	52.2	786	3	AAB07739 A snake v
8	2102.5	52.2	787	3	AAB07740 A snake v
9	2102.5	52.2	820	3	AAB07741 A snake v
10	2102.5	52.2	820	7	ADJ69446 Human hea
11	1717.5	42.6	542	2	AAY03224 Amino aci
12	1654	41.0	535	4	AEL3056 ADAM-20di
13	1646	40.8	535	2	AAY17414 SVPH1-26
14	1387.5	34.4	790	3	AAB07705 Human aci
15	1387.5	34.4	790	4	AAU12273 Human PRO
16	1387.5	34.4	790	6	ABO17717 Novel hum
17	1387.5	34.4	790	6	ABU80971 Human PRO
18	1387.5	34.4	790	6	ABU66671 Human PRO
19	1387.5	34.4	790	6	ABU59752 Novel sec
20	1387.5	34.4	790	6	ABO24942 Human sec
21	1387.5	34.4	790	6	ABU66947 Human sec
22	1387.5	34.4	790	6	ADA45723 Novel hum
23	1387.5	34.4	790	6	ADA76154 Human PRO
24	1387.5	34.4	790	6	ADA18804 Human PRO
25	1387.5	34.4	790	6	ADA61427 Homo sapi

26	1387.5	34.4	790	6	ADB19212 Novel hum
27	1387.5	34.4	790	6	ADB27753 Human PRO
28	1387.5	34.4	790	6	ADA86232 Novel hum
29	1387.5	34.4	790	6	ADB15796 Human PRO
30	1387.5	34.4	790	6	ADA47582 Human PRO
31	1387.5	34.4	790	6	ADA67377 Human PRO
32	1387.5	34.4	790	6	ADB30384 Human PRO
33	1387.5	34.4	790	6	ADA85680 Novel hum
34	1387.5	34.4	790	6	ADA96892 Human PRO
35	1387.5	34.4	790	6	ADA79196 Human PRO
36	1387.5	34.4	790	6	ADA87335 Novel hum
37	1387.5	34.4	790	6	ADB16537 Human PRO
38	1387.5	34.4	790	6	ADA91629 Novel hum
39	1387.5	34.4	790	6	ADB14692 Human PRO
40	1387.5	34.4	790	6	ADB18653 Novel hum
41	1387.5	34.4	790	6	ADA93868 Human PRO
42	1387.5	34.4	790	6	ADB19764 Novel hum
43	1387.5	34.4	790	6	ADB13076 Human PRO
44	1387.5	34.4	790	6	ABO43250 Novel hum
45	1387.5	34.4	790	6	ADA74330 Human PRO

#### ALIGNMENTS

RESULT 1

AAY17413

ID AAY17413 standard; protein; 726 AA.

XX

AC AAY17413;

DT 26-JUL-1999 (first entry)

XX

DE Human SVPH1-26 protein.

XX

KW Human; SVPH1-26; proteinase; testis; fertilisation; spermatogenesis;

KW birth control; detergent additive; diagnosis; testicular cancer.

XX

OS Homo sapiens.

XX

PN WO9923228-A1.

XX

PD 14-MAY-1999.

XX

PF 30-OCT-1998; 98WO-US022965.

XX

PR 30-OCT-1997; 97US-0063571P.

XX

PA (IMMV ) IMMUNEX CORP.

XX

PI Cerretti DP;

XX

DR WPI; 1999-337487/28.

XX

DR N-PSDB; AAX56461.

XX

PT New human SVPH1-26 DNA useful for the diagnosis and prognosis of

XX

PT testicular cancers.

XX

PS Claim 2; Page 73-83; 96pp; English.

XX

CC The present sequence is human SVPH1-26. Human SVPH1-26 polypeptides are  
 CC proteinases implicated in fertilisation and spermatogenesis. They can be  
 CC used as therapeutic agents. A proteinase inhibitor of the catalytic  
 CC domain would inhibit SVPH1-26 activity and would be useful as a method  
 CC for birth control. Also, an inhibitor of the integrin domain of SVPH1-  
 CC 26 may affect fertilisation. The proteinase activity of SVPH1-26 can also  
 CC be used as a detergent additive for the removal of stains having a  
 CC protein component. The SVPH1-26 polypeptides and fragments can also be  
 CC used as molecular weight markers, as markers for determination of  
 CC isoelectric points of sample proteins and as controls for establishing  
 CC the extent of fragmentation of a protein sample. The products can also be  
 CC used for identifying, separating or purifying cells that express SVPH1-26  
 CC polypeptides such as testis cells. They can be used for the diagnosis and





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 16:31:09 ; Search time 124 Seconds  
(without alignments)  
2998.141 Million cell updates/sec

Title: US-10-633-202-2

Perfect score: 4030

Sequence: 1 NAVGSLVHVRVTLMLWFG.....LFCLHVLFKKRTKSKDEEG 726

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query %	Length	DB	ID	Description
1	4004	99.4	726	1	AD20	HUMAN	O43506 homo sapien
2	2435	60.4	722	1	AD21	HUMAN	O9ukj8 homo sapien
3	2102.5	52.2	820	1	AD29	HUMAN	O9ukf5 homo sapien
4	2050	50.9	729	1	AD21	MOUSE	O9ji76 mus musculus
5	1994.5	49.5	760	1	AD35	MOUSE	O9ri59 mus musculus
6	1953.5	48.5	756	2	Q7M762		Q7m762 mus musculus
7	1946	48.3	760	2	Q7M763		Q7m763 mus musculus
8	1934	48.0	763	2	Q811Q4		Q811q4 mus musculus
9	1899.5	47.1	761	2	Q8CDV3		Q8cdv3 mus musculus
10	1892.5	47.0	761	1	AD24	MOUSE	Q9ri60 mus musculus
11	1780.5	44.2	714	2	Q8K4K0		Q8k4k0 mus musculus
12	1774	44.0	713	2	Q7M765		Q7m765 mus musculus
13	1772.5	44.0	702	2	Q8BMR5		Q8bmr5 mus musculus
14	1759.5	43.7	736	2	Q7M766		Q7m766 mus musculus
15	1758	43.6	699	2	Q6IMH6		Q6imh6 mus musculus
16	1707.5	42.4	697	2	Q8BMR4		Q8bmr4 mus musculus
17	1697.5	42.1	697	1	AD26	MOUSE	Q9ri58 mus musculus
18	1387.5	34.4	790	1	AD30	HUMAN	O9ukf2 homo sapien
19	1387.5	34.4	790	2	Q8TBZ7		Q8tbz7 homo sapien
20	1338.5	33.2	819	1	AD09	HUMAN	Q13443 homo sapien
21	1333.5	33.1	775	2	Q76Kt5		Q76kt5 gallus gall
22	1332	33.1	873	2	Q42595		Q42595 xenopus lae
23	1328.5	33.0	845	1	AD09	MOUSE	Q61072 mus musculus
24	1300.5	32.3	732	2	Q811Q3		Q811q3 mus musculus
25	1250.5	31.0	838	2	O19056		O19056 papio anubi
26	1237.5	30.7	905	2	Q28476		Q28476 macaca fasc
27	1231	30.5	825	2	O46652		O46652 papio anubi
28	1230.5	30.5	821	2	O19060		O19060 saguinus oe
29	1229.5	30.5	836	2	O19057		O19057 pongo pygma
30	1226	30.4	825	2	Q28477		Q28477 macaca fasc
31	1215	30.1	792	2	O19061		O19061 saguinus oe

#### RESULT 1

ID	AD20	HUMAN	STANDARD;	PRT;	726	AA.
AC	O43506;	O9UKJ9;				
DT	16-OCT-2001	(Rel. 40, Created)				
DT	16-OCT-2001	(Rel. 40, Last sequence update)				
DT	25-OCT-2004	(Rel. 45, Last annotation update)				
DE	ADAM 20 precursor	(EC 3.4.24.-) (A disintegrin and metalloproteinase domain 20).				
DE	domain 20.					
GN	Name=ADAM20;					
OS	Homo sapiens	(Human).				
OC	Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria;	Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Testis;					
RX	MEDLINE=98137801;	PubMed=9469942; DOI=10.1016/S0378-1119(97)00597-0;				
RA	Hoofst van Huijsduijn R.;					
RT	"ADAM 20 and 21; two novel human testis-specific membrane metalloproteases with similarity to fertilin-alpha.";					
RL	Gene 206:273-282 (1998).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=99453762;	PubMed=10524237; DOI=10.1016/S0378-1119(99)00302-9;				
RA	Poindexter K., Nelson N., DuBose R.F., Black R.A., Cerretti D.P.;					
RT	"The identification of seven metalloproteinase-disintegrin (ADAM) genes from genomic libraries.";					
RL	Gene 237:61-70 (1999).					
CC	-!- FUNCTION: May be involved in sperm maturation and/or fertilization.					
CC	-!- COPACTOR: Binds 1 zinc ion per subunit (Potential).					
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.					
CC	-!- TISSUE SPECIFICITY: Testis specific.					
CC	-!- DOMAIN: A tripeptide motif (VGE) within disintegrin-like domain could mediate sperm/egg binding.					
CC	-!- DOMAIN: The cysteine-rich domain encodes putative cell-fusion peptides, which could be involved in sperm-egg fusion.					
CC	-!- PTM: Has no obvious cleavage site for furin endopeptidase, suggesting that the proteolytic processing is regulated.					
CC	-!- MISCELLANEOUS: May be the functional equivalent of ADAM 1/fertilin alpha which is a pseudogene in human.					
CC	-!- SIMILARITY: Belongs to the peptidase M12B family.					
CC	-!- SIMILARITY: Contains 1 disintegrin domain.					
CC	-!- SIMILARITY: Contains 1 EGF-like domain.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ) or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .					

#### ALIGNMENTS

	32	1210.5	30.0	789	2	Q66HK9	Q66hk9	rattus norv
33	1208.5	30.0	789	1	AD01	RAT	P70505	rattus norv
34	1204	29.9	791	1	AD1A	MOUSE	Q60813	mus musculus
35	1193	29.6	812	2	O77779		O77779	bos taurus
36	1193	29.6	919	2	O28659		O28659	oryctolagus
37	1190.5	29.5	763	2	Q8CGQ2		Q8csg2	mus musculus
38	1170.5	29.0	713	2	Q28485		Q28485	macaca fasc
39	1169.5	29.0	703	2	Q8C0V3		Q8cdv3	mus musculus
40	1168.5	28.9	756	2	O6IMH7		Q6imh7	mus musculus
41	1163.5	28.9	730	2	O19051		O19051	oryctolagus
42	1160.5	28.8	754	2	O8CGQ1		Q8csg1	mus musculus
43	1153.5	28.6	732	2	Q28484		Q28484	macaca fasc
44	1152.5	28.6	751	2	P70535		P70535	rattus norv
45	1149	28.5	751	2	Q6IMH8		Q6imh8	mus musculus

CC DR ENBL; AF029899; AAC52041.1; --  
DR EMBL; AF158643; AAD52524.1; --  
DR HSSP; P18619; IFVL.  
DR MEROPS; M12.218; --  
DR MTM; 603712; --  
DR GO; GO:0008237; F-metalloproteinase activity; TAS.  
DR GO; GO:0007338; P-fertilization (sensu Animalia); TAS.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR001818; Pept M10A\_M12B.  
DR InterPro; IPR006025; Pept M Zn BS.  
DR InterPro; IPR001590; Peptidase\_M12B.  
DR InterPro; IPR002870; Peptidase\_M12B\_N.  
DR Pfam; PF00200; Disintegrin; 1.  
DR Pfam; PF01562; Pep M12B propep; 1.  
DR Pfam; PF01421; Reprolysin; 1.  
DR PRINTS; PR00289; DISINTEGRIN.  
DR ProDom; PD000664; Disintegrin; 1.  
DR PROSITE; PS02015; ADAM\_MPRO; 1.  
DR PROSITE; PS00546; CYSTEINE\_SWITCH; FALSE NEG.  
DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
DR PROSITE; PS02014; DISINTEGRIN\_2; 1.  
DR PROSITE; PS00022; EGF\_1; FALSE NEG.  
DR PROSITE; PS01186; EGF\_2; FALSE NEG.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW EGF-like domain; Glycoprotein; Hydrolase; Metalloprotease; Signal;  
KW Transmembrane; Zinc; Zymogen.  
FT SIGNAL 1 31 Potential.  
FT PROPEP 32 206 Potential.  
FT CHAIN 207 726 ADAM 20.  
FT DOMAIN 207 693 Extracellular (Potential).  
FT TRANSMEM 694 714 Potential.  
FT DOMAIN 715 726 Cytoplasmic (Potential).  
FT DOMAIN 207 384 Metalloprotease.  
FT DOMAIN 407 493 Disintegrin-like.  
FT DOMAIN 494 634 Cys-rich.  
FT DOMAIN 635 663 EGF-like.  
FT SITE 173 173 Cysteine switch.  
FT ACT\_SITE 343 343 Potential.  
FT METAL 342 342 Zinc (catalytic) (Potential).  
FT METAL 346 346 Zinc (catalytic) (Potential).  
FT METAL 352 352 Zinc (catalytic) (Potential).  
FT DISULFID 317 394 By similarity.  
FT DISULFID 357 379 By similarity.  
FT DISULFID 465 478 Potential.  
FT DISULFID 635 646 By similarity.  
FT DISULFID 640 652 By similarity.  
FT DISULFID 654 663 By similarity.  
FT CARBOHYD 191 191 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 226 226 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 378 378 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 438 438 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 479 479 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 587 587 N-linked (GlcNAc... ) (Potential).  
FT CONFLICT 19 19 L -> F (in Ref. 2).  
FT CONFLICT 109 109 R -> C (in Ref. 2).  
FT CONFLICT 637 637 R -> P (in Ref. 2).  
SQ SEQUENCE 726 AA; 81710 MW; BF356A6771CD1C12 CRC64;  
  
Query Match 99.4%; Score 4004; DB 1; Length 726;  
Best Local Similarity 99.6%; Pred. No. 1.7e-281;  
Matches 723; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 MAVGEPLVHVRVTLVLLWFGNFIISGHSQARPSQFTSPSEVVIPLKVISRGGAAPGW 60  
Db 1 MAVGEPLVHVRVTLVLLWFGNFIISGHSQARPSQFTSPSEVVIPLKVISRGGAAPGW 60  
  
Qy 61 LSYSLRFGGORYIVHMRVKNLLFAAHLPVFTYEQHALLQDQPFIDDCYHYGVGVPE 120  
Db 61 LSYSLRFGGORYIVHMRVKNLLFAAHLPVFTYEQHALLQDQPFIDDCYHYGVGVPE 120

Qy 121 SLVALSTCGSGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDDTQFPFMRGCLTEEKI 180  
Db 121 SLVALSTCGSGFLGMLQINDLVYEIKPISVSATFEHLVYKIDSDDTQFPFMRGCLTEEKI 180  
  
Qy 181 AHOMELQLSYNFTLKOSSVFGWTHORFVELVVDNIRYLFSSQSNATTVOHEVFVNVI 240  
Db 181 AHOMELQLSYNFTLKOSSVFGWTHORFVELVVDNIRYLFSSQSNATTVOHEVFVNVI 240  
  
Qy 241 VDSFYHPLVDVLLTGIDITWASNPLPTSGDDLNDVLEDFSIWKNNLNRLQHDVAHLFI 300  
Db 241 VDSFYHPLVDVLLTGIDITWASNPLPTSGDDLNDVLEDFSIWKNNLNRLQHDVAHLFI 300  
  
Qy 301 KDTQGMKLGVAVYKIGICQNPFTGVDVFDNRLVWFATLGHGHLNLMQHDTCVCE 360  
Db 301 KDTQGMKLGVAVYKIGICQNPFTGVDVFDNRLVWFATLGHGHLNLMQHDTCVCE 360  
  
Qy 361 LOWCIMHAYRVKVTTFKPSNCSYAYQWDSITSSGLCIQPPYPGNIPLKCYCGNLVVEGEE 420  
Db 361 LOWCIMHAYRVKVTTFKPSNCSYAYQWDSITSSGLCIQPPYPGNIPLKCYCGNLVVEGEE 420  
  
Qy 421 CDCGTIROCADPCCLLNCTLHPGACAFGICCKCKFLPSGTLCRQQVGECDLPWCNG 480  
Db 421 CDCGTIROCADPCCLLNCTLHPGACAFGICCKCKFLPSGTLCRQQVGECDLPWCNG 480  
  
Qy 481 TSHQCPDDVYVQDGISSCNVNAFCYEKTCNNHDIQCKEIEFGQDARSASQCYQINTQGNR 540  
Db 481 TSHQCPDDVYVQDGISSCNVNAFCYEKTCNNHDIQCKEIEFGQDARSASQCYQINTQGNR 540  
  
Qy 541 FGHGIGIVGTYVKWTPDTCGRVQCVNPGVTPNLIEHSTVQGFHLNDDTCWGTDTYHLGM 600  
Db 541 FGHGIGIVGTYVKWTPDTCGRVQCVNPGVTPNLIEHSTVQGFHLNDDTCWGTDTYHLGM 600  
  
Qy 601 ATPDIGEVKDGTVCGPEKICIRKCKASVHLVSOACOPKTCNMRGICNNKQHCNHEWAP 660  
Db 601 ATPDIGEVKDGTVCGPEKICIRKCKASVHLVSOACOPKTCNMRGICNNKQHCNHEWAP 660  
  
Qy 661 PYCKDKGYGGSADSGPPPKNNMEGLNVGMGLRYLSLLCLLPLVAFLLFCLHVLFKKRTKS 720  
Db 661 PYCKDKGYGGSADSGPPPKNNMEGLNVGMGLRYLSLLCLLPLVAFLLFCLHVLFKKRTKS 720  
  
Qy 721 KEDEEG 726  
Db 721 KEDEEG 726  
  
RESULT 2  
AD21\_HUMAN STANDARD; PRT; 722 AA.  
ID AD21\_HUMAN STANDARD; PRT; 722 AA.  
AC Q9UKJ8; O43507;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE ADAM 21 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 21).  
GN Name=ADAM21;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99453762; PubMed=10524237; DOI=10.1016/S0378-1119(99)00302-9;  
RA Poindexter K., Nelson N., DuBoise R.F., Black R.A., Cerretti D.P.;  
RT "The identification of seven metalloproteinase-disintegrin (ADAM) genes from genomic libraries."  
RL Gene 237:61-70(1999).  
RN [2]  
RP SEQUENCE OF 181-722 FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=98137801; PubMed=9469942; DOI=10.1016/S0378-1119(97)00597-0;  
RA Hoof van Huifeduijnen R.;  
RT "ADAM 20 and 21; two novel human testis-specific membrane metalloproteases with similarity to fertilin-alpha.";



